République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
École Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès

THESIS

To obtain the diploma of **Master**
Field: **Computer Science**
Specialty: **Artificial Intelligence and Data Sciences (IASD)**

Theme

Deep Learning Approaches for Multimodal Fake News Detection
Text and Image Perspectives

Presented by:
Boudali Riadh

Submission Date: **September, 2025**
In front of the jury composed of:

Dr. KHALDI Belkacem	President
Dr. Serhane Oussama	Supervisor
Dr. NEGGAZ Imene	Examiner

Academic Year: 2024/2025

0.1 Abstract

The rapid proliferation of misinformation on social media has made automated detection a critical research priority. This thesis presents a focused, systematic study of deep learning approaches for multimodal fake-news detection, emphasizing the joint use of textual and visual signals. We review and categorize methods across three axes: text-based techniques (including transformer hybrids), image-based forensic and deep models (CNNs, Vision Transformers), and multimodal frameworks that fuse or explicitly assess cross-modal consistency (e.g., adversarial event-invariant models, MVAE, SAFE, and recent CLIP-guided and multi-scale attention systems). Through comparative analysis of datasets, architectures, and evaluation protocols, we identify key strengths — notably the semantic alignment offered by pretrained vision–language models — and persistent weaknesses, including dataset scarcity, modality imbalance, domain shift, computational cost, and limited explainability. Our contribution is a consolidated taxonomy, critical synthesis of results, and a set of recommended directions for future work: (i) designing data-efficient, domain-adaptive multimodal models, (ii) improving interpretability for real-world deployment, and (iii) developing robust defenses against evolving adversarial misinformation. The thesis demonstrates that combining cross-modal semantic reasoning with pretrained vision–language backbones yields the most promising path toward robust, generalizable fake-news detection.

Keywords:

fake news detection, multimodal learning, transformers, CLIP, semantic alignment

0.2 Résumé

La diffusion rapide de la désinformation sur les réseaux sociaux exige des approches automatiques performantes. Ce mémoire propose une étude systématique des approches d'apprentissage profond pour la détection multimodale de fake news, centrée sur l'analyse conjointe du texte et de l'image. Nous présentons une taxonomie des méthodes : approches textuelles (notamment les modèles Transformer et leurs hybrides), approches visuelles (méthodes forensiques, CNN, Vision Transformers) et méthodes multimodales qui fusionnent les modalités ou évaluent explicitement la cohérence sémantique (par ex. EANN, MVAE, SAFE, puis des modèles récents guidés par CLIP ou par attention multi-échelle). Par une comparaison critique des jeux de données, architectures et protocoles d'évaluation, nous mettons en évidence les progrès (alignement sémantique via modèles vision-language préentraînés) ainsi que les limites persistantes : faiblesse des jeux de données, déséquilibre de modalités, décalage de domaine, coût computationnel et manque d'explicabilité. Les contributions incluent une synthèse critique, une classification structurée des approches et des pistes futures : (i) modèles multimodaux économes en données et adaptatifs au domaine, (ii) mécanismes d'explicabilité pour l'utilisation réelle, (iii) stratégies robustes face aux menaces adversariales évolutives. L'étude conclut que l'intégration de l'alignement sémantique inter-modal avec des backbones vision-language constitue la direction la plus prometteuse.

Mots-clés : détection de fake news, multimodal, transformeurs, CLIP, alignement sémantique.